

AAAAATAAATCAT ATG AAA AAT ATT AAA AAA AAT CAA GTA ATC AAT CTC GGT CCT AAT TCT
 M K N I K K N Q V M N L G P N S
 AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA AAT ATT GAA CAA TTT GAA GCA
 K L L K E Y K S Q L I E L N I E Q F E A
 GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT
 G I G L I L L G D A Y I R S R D E G K T Y
 TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GTA TGT TTA TTA TAT GAT
 C M Q F E W K N K A Y M D H V C L L Y D
 CAA TGG GTA TTA TCA CCT CAT AAA AAA GAA AGA GTT AAT CAT TTA GGT AAT TTA GTA
 Q W V L S P P H K K E R V N H L G N L V
 ATT ACC TGG GGA GCT CAA ACT TTT AAA CAT CAA GCT TTT AAT AAA TTA GCT AAC TTA TTT
 I T W G A Q T F K H Q A F N K L A N L F
 ATT GTA AAT AAT AAA CTT ATT CTT AAT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG
 I Y N N K K L I P N N L V E N Y L T P M
 AGT CTG GCA TAT TGG TTT ATG GAT GAT GGA GGT AAA TGG GAT TAT AAT AAA AAT TCT CTT
 S L A Y W F M D D G G K W D Y N K N S L
 AAT AAA AGT ATT GTA TTA AAT ACA CAA AGT TTT ACT TTT GAA GAA GTA GAA TAT TTA CTT
 N K S I V L L N T Q S F T F E E V C Y L V
 AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT TAT GTT AAA ATT AAT AAA AAT AAA CCA
 K G L R N K F Q L N C Y V K I N K N K P
 ATT ATT TAT ATT GAT TCT AGT AGT TAT CTG ATT TTT TAT AAT TTA ATT AAA CCT TAT TTA
 I I Y I D S M S Y L I F Y N I T K P Y L
 ATT CCT CAA ATG ATG TAT AAA CTG CCT AAT ACT ATT TCA TCC GAA ACT TTT TTA AAA TAA
 I P Q M Y K L P N T I S S E T F L K

FIG. 1

Bam HI

1. CCGGATCCATG CAT ATG AAA AAC ATC AAA AAC CAG GTA ATG AAC CTG GGT CCG AAC TCT
 AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC ATC GAA CAG TTC GAA GCA
 GGT ACT GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT GAT GAA GGT AAA ACC TAC
 TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC GTA TGT CTG CTG TAC GAT
 CAG TGG GTA CTG TCC CCG CCG CAC AAA AAC GAA CGT GTT AAC CAC CTG GGT AAC CTG GTA
 ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC AAA CTG GCT AAC CTG TTC
 ATC GTT AAC AAC AAA ACC ATC CCG AAC AAC CTG GTT GAA AAC TAC CTG ACC CCG ATG
 TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT TAC AAC AAC TCT ACC
 AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA GTA GAA TAC CTG GTT
 AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC CTA AAA ATC AAC AAA AAC AAA CCG
 ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC CTG ATC AAA CCG TAC CTG
 ATC CCG CAG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC GAA ACT TTC CTG AAA TAA
 TAAGTCGACTGCAGGATCCGGTAAGTAAGTAA
2. Sall PstI BamHI

1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC
 ACTIVITY. OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS
 OF THE 10 FIRST AMINO ACIDS.

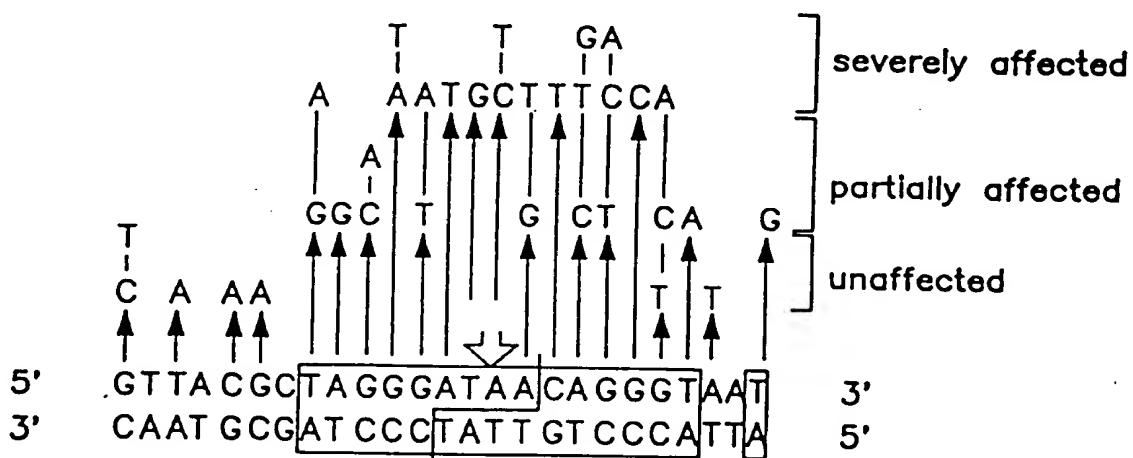


FIG. 3

1667 GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGCGCAGGAGGCTTCCAGGGGGAACGCCCTGGTATCT 1746
 1747 TTATAGTCCIGTCGGGTTTCGCCACCCTCIGACTTGAGCGTCGATTTTGTG ATG CTC GTC AGG GGG GCG GAG 1818
 1819 CCT ATG GAA AAA CGC CAG CAA CGC GGC CTT TTT ACG GTT CCT GGC CTT TTG CTG GCC TTT 1878
 1879 TGC TCA CAT GTT CTT ICC TGC GTT ATC CCC TGA TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC 1947
 1948 TGATACCGCTCGCCGCAGCCGACGAGCGAGCGAGGAGGAGCGGAGGAGCGCCCAATACGCAAC 2027
 2028 CGCCTCTCCCCGCGGTTGGCCGATTCAATTA ATG CAG CTG GCA CAG GTT TCC CGA CTG GAA AGC 2094
 2095 GGG CAG TGA GCGCAACGCAATTA ATG TGA GTTAGCTACCTCATTAGGCACCCCGAGGCTTTACACTTT ATG 2164
 2165 CTT CCG GCT CGT ATG TTG TGT GGA ATT GTG AGC GGA TAA CAATTTACACAGGAACAGCT ATG 2228
 2229 ACC ATG ATT ACG AAT TCT CAT GTT TGA CAGCTTATCATCGATAAGCTTTA ATG CGG TAG TTTATCAC 2295
 2296 AGTTAAATTGCTAACGCAGTCAGGCACCGTGT ATG AAA TCT AAC AAT GCG CTC ATC GTC ATC CTC GGC 2363
 2364 ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA CTG CCG GGC CTC TTG 2423
 2424 CGG GAT ATC CGC CTG ATG CGT GAA CGT GAC GGA CGT AAC CAC CGC GAC ATG TGT GTG CTG 2483
 2484 TTC CGC TGG GCA TGC CAG GAC AAC TTC TGG TCC GGT AAC GTG CTG AGC CCG GCC AAG CTT 2543

FIG. 4A

2544	ACT	CCC	CAT	CCC	CCT	GTT	GAC	AAT	TAA	TCATCGGCTCGTATA	ATG	TGT	GGA	ATT	GTG	AGC	GGA	2606
73	T	P	H	P	P	V	D	N	*				G	I	V	S	G	7
2607	TAA	CAATTT	CACACAGGAA	CAGGATCC														8
2671	AAC	CTG	GGT	CCG	AAC	TCT	AAA	CTG	CTG	AAA	GAA	TAC	AAA	TCC	CAG	CTG	ATC	2670
13	N	L	G	P	P	N	S	K	L	L	K	E	Y	K	S	Q	L	12
2731	ATC	GAA	CAG	TTC	GAA	GCA	GGT	ATC	GGT	CTG	ATC	CTG	GAT	GCT	TAC	ATC	CTG	2730
33	I	E	Q	F	E	A	G	I	G	L	I	L	G	D	A	Y	I	32
2791	GAT	GAA	GGT	AAA	ACC	TAC	TGT	ATG	CAG	TTC	GAG	TGG	AAA	AAC	AAA	GCA	TAC	2790
53	D	E	G	K	T	Y	C	M	Q	F	E	W	K	N	K	A	Y	52
2851	GTA	TGT	CTG	CTG	TAC	GAT	CAG	TGG	GTA	CTG	TCC	CCG	CCG	CAC	AAA	AAA	GAA	2850
73	V	C	L	L	L	Y	D	Q	W	V	L	S	P	H	K	K	E	72
2911	CAC	CTG	GGT	AAC	CTG	GTA	ATC	ACC	TGG	GGC	GCC	CAG	ACT	TTC	AAA	CAC	CAA	2910
93	H	L	G	N	L	V	I	T	W	G	A	Q	T	F	K	H	Q	92
2971	AAA	CTG	GCT	AAC	CTG	TTC	ATC	GTT	AAC	AAC	AAA	AAA	ACC	ATC	CCG	AAC	AAC	2970
113	K	L	A	N	L	F	I	V	N	N	K	K	T	I	P	N	N	112
3031	AAC	TAC	CTG	ACC	CCG	ATG	TCT	CTG	GCA	TAC	TGG	TTC	ATG	GAT	GGT	GGT	AAA	3030
133	N	Y	L	T	P	M	S	L	A	Y	W	F	M	D	G	G	K	132
3091	TAC	AAC	AAA	AAC	TCT	ACC	AAC	AAA	TCG	ATC	GTA	CTG	AAC	ACC	CAG	TCT	TTC	3090
153	Y	N	K	N	S	T	N	K	S	I	V	L	N	T	Q	S	F	152
3151	GAA	GTA	GAA	TAC	CTG	GTT	AAG	GGT	CTG	CGT	AAC	AAA	TTC	CAA	CTG	AAC	TGT	3150
173	E	V	E	Y	L	V	K	G	L	R	N	K	F	Q	L	N	C	172
3211	ATC	AAC	AAA	AAC	AAA	CCG	ATC	ATC	TAC	ATC	GAT	TCT	ATG	TCT	TAC	CTG	ATC	3210
193	I	N	K	N	K	P	I	I	Y	I	D	S	M	S	Y	L	I	192
3271	CTG	ATC	AAA	CCG	TAC	CTG	ATC	CCG	CAG	ATG	ATG	TAC	AAA	CTG	CCG	AAC	ACT	3270
213	L	I	K	P	Y	L	I	P	Q	M	M	Y	K	L	P	N	T	212
3331	GAA	ACT	TTC	CTG	AAA	TAA												3330
233	E	T	F	L	K													232

FIG. 4B

		-2	-1	1				5					10						
		M	H	M	K	N	I	K	K	N	Q	V	M	N	L	G	P	N	S
K	L	<u>L</u>	20 K	E	Y	K	S	Q	L	I	E	L	30 N	I	E	Q	F	E	<u>A</u>
G	<u>I</u>	<u>G</u>	40 <u>L</u>	<u>I</u>	<u>L</u>	<u>G</u>	<u>D</u>	<u>A</u>	<u>Y</u>	<u>I</u>	R	S	50 R	D	E	G	K	T	Y
C	M	Q	60 F	E	W	K	N	K	A	Y	M	D	70 H	V	C	L	L	Y	C
Q	W	Y	80 <u>L</u>	S	P	P	H	K	K	E	R	Y	90 N	<u>H</u>	L	G	N	L	Y
I	T	W	100 G	A	Q	T	F	K	H	Q	A	F	110 N	K	L	A	N	L	F
I	V	N	120 N	K	K	<u>I</u>	I	P	N	N	L	V	130 E	N	Y	L	T	P	M
G	L	A	140 Y	W	P	M	<u>D</u>	<u>D</u>	<u>G</u>	<u>G</u>	K	W	150 D	Y	N	K	N	S	<u>I</u>
N	K	S	160 I	V	L	N	T	Q	S	F	T	F	170 E	E	V	E	Y	L	V
K	G	L	180 R	N	K	F	Q	L	N	C	Y	V	190 K	I	N	K	N	K	P
I	I	Y	200 I	D	S	M	S	Y	L	I	F	Y	210 N	L	I	K	P	Y	L
I	P	Q	220 M	M	Y	K	L	P	N	T	I	S	230 S•	E	T	F	L	K	•

```
positions 1 to 10: can be deleted
position 36: G is tolerated
position 40: M or V are tolerated
position 41: S or N are tolerated
position 43: A is tolerated
position 46: V or N are tolerated
position 91: A is tolerated
positions 123 and 156: L are tolerated
position 223: A and S are tolerated
```

position 19: L to S
 position 38: L to S or N
 position 39: G to D or R
 position 40: L to Q
 position 42: L to R
 position 44: D to E G or H
 position 45: A to E or D
 position 46: Y to D
 position 47: Y to R or N
 position 80: L to S
 position 144: D to E
 position 145: D to E
 position 146: G to E
 position 147: G to S

FIG. 5

Group I Intron Encoded Endonucleases and Related Endonucleases

ENDONUCLEASE		RECOGNITION SEQUENCE	CLEAVAGE SITE	▽ INTRON SITE
TWO DODECAPEPTIDE FAMILY (OR 4 BP CUTTERS)	I-Sce I (Saccharomyces mitochondria)	CGCTAGGGATAACAGGGTAATATAGC GCGATCCCTATTGTCCCATTATATCG		
	I-Sce IV (Saccharomyces mitochondria)	TTCTCATGATTAGCTCTAATCCATGG AAGAGTACTAATCGAGATTAGGTACC		
	I-Sce II (Saccharomyces mitochondria)	CTTTGGTCAATCCAGAAGTATATATTT GAAACCAGTAGGTCTTCATATATAAAA		
	I-Ceu I (Chlamydomonas chloroplast)	TAA CGGTCTTAAAGGTAGCGAAATTCA ATTGCCAGGATTC CATCGCTTTAAGT		
	I-Ppo I (Physarum nucleus)	TGACTCTCTTAAAGGTAGCCAAATGCC ACTGAGAGGAATTC CATCGGTTTACGG		
	I-Sce III (Saccharomyces mitochondria)	GGAGGTTTTTGGTAACTATTTATTACC CCTCCAAAACCATTGATAAATAATGG		
	I-Cre I (Chlamydomonas chloroplast)	GGGTTCAAAACGTCTGTGAGACAGTTT CCCAAGTTTTTG CAGCACTCTGTCAA		
	Endo. Sce I(RF3) (Saccharomyces mitochondria) (Non intronic)	GATGCTGTAGGCATAGGCTTGTTTAT CTACGACATCCGTATCCGAACCAATA		
	HO (Saccharomyces nucleus) (Non intronic)	CTTTCCGCAACAGTATAATTTTATAA GAAAGGCGTTGT CATATTTAAAATATT		
	I-Csm I (Chlamydomonas mitochondria) (Putative endonuclease)	ACCATGGGGTCAAATGTCTTTCTGGG TGGTACCCAGTTTACAGAAAGACCC		
OTHER STRUCTURAL FAMILIES	I-Pan I (Podospira mitochondria) (Putative endonuclease)	GTGCCTGAATGATATTTATTACCTTT CACGGACTTACTATAAATAATGGAAA		
	(Bacteriophage T4)			
	I Tev I	CAACGCTCAGTAGATGTTTTCTTGGGTCTACCGTTTAAAT GTTGCGAGTCATCTACAAAAGAACCCAGATGGCAAATTA		
	I Tev II	CAAGCTTATGAGTATGAAGTGAACACGTTATT GTTCGAATACTCATACTTCACTTGTGCAATAA		
	I Tev III	GCTATTGTTTTTATGATCTTTTGCCTGTAGCTTTAA CGATAAGCAAAAATA CATAGAAAACGCACATCGAAATT		

FIG. 6

EXPRESSION VECTORS

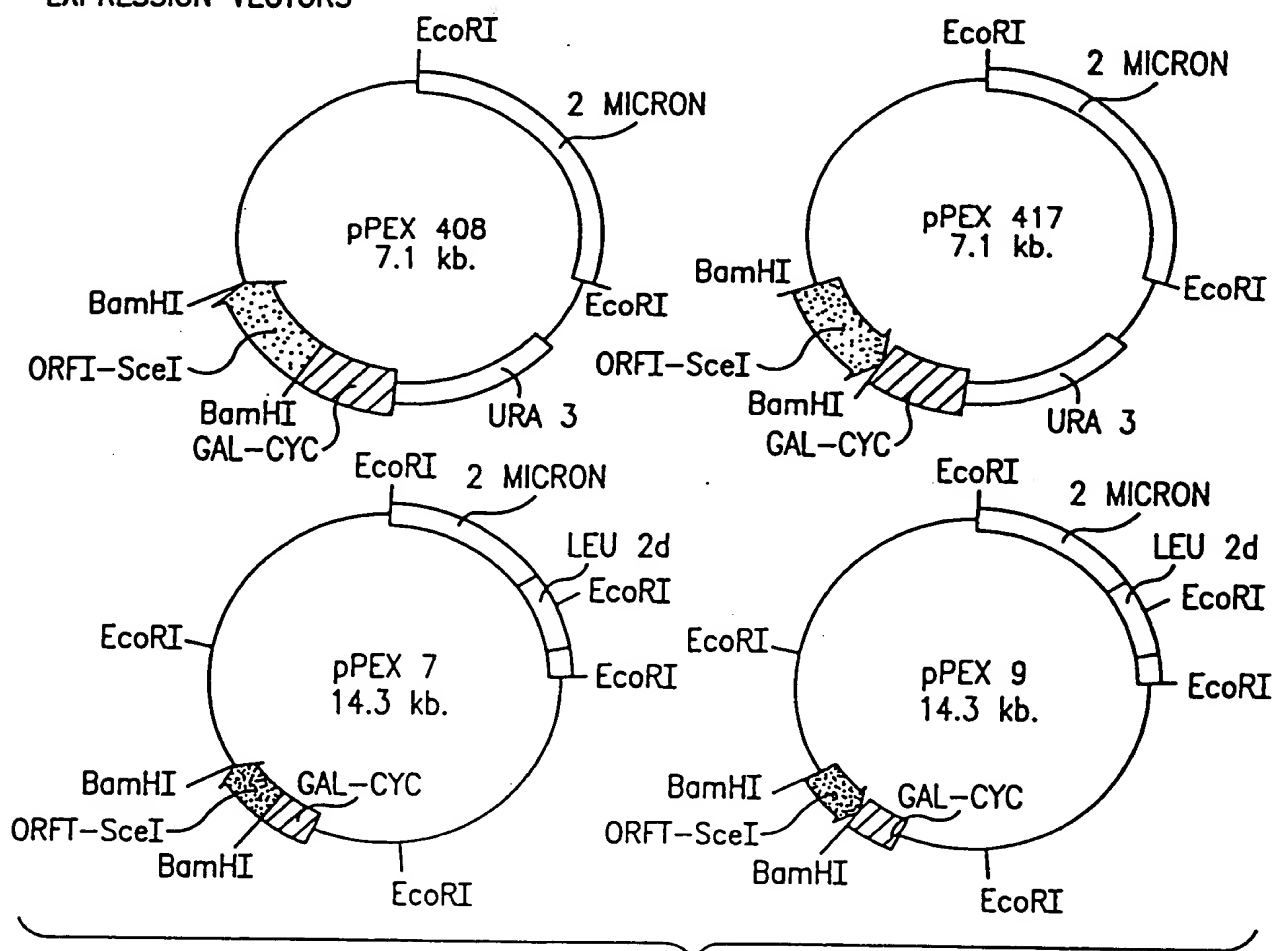


FIG. 7

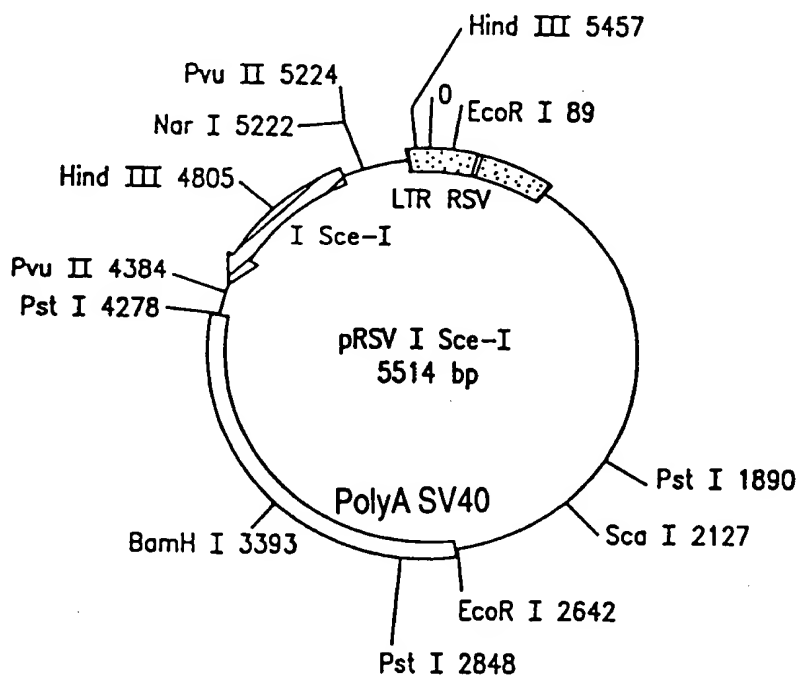


FIG. 8

003416 0919E860

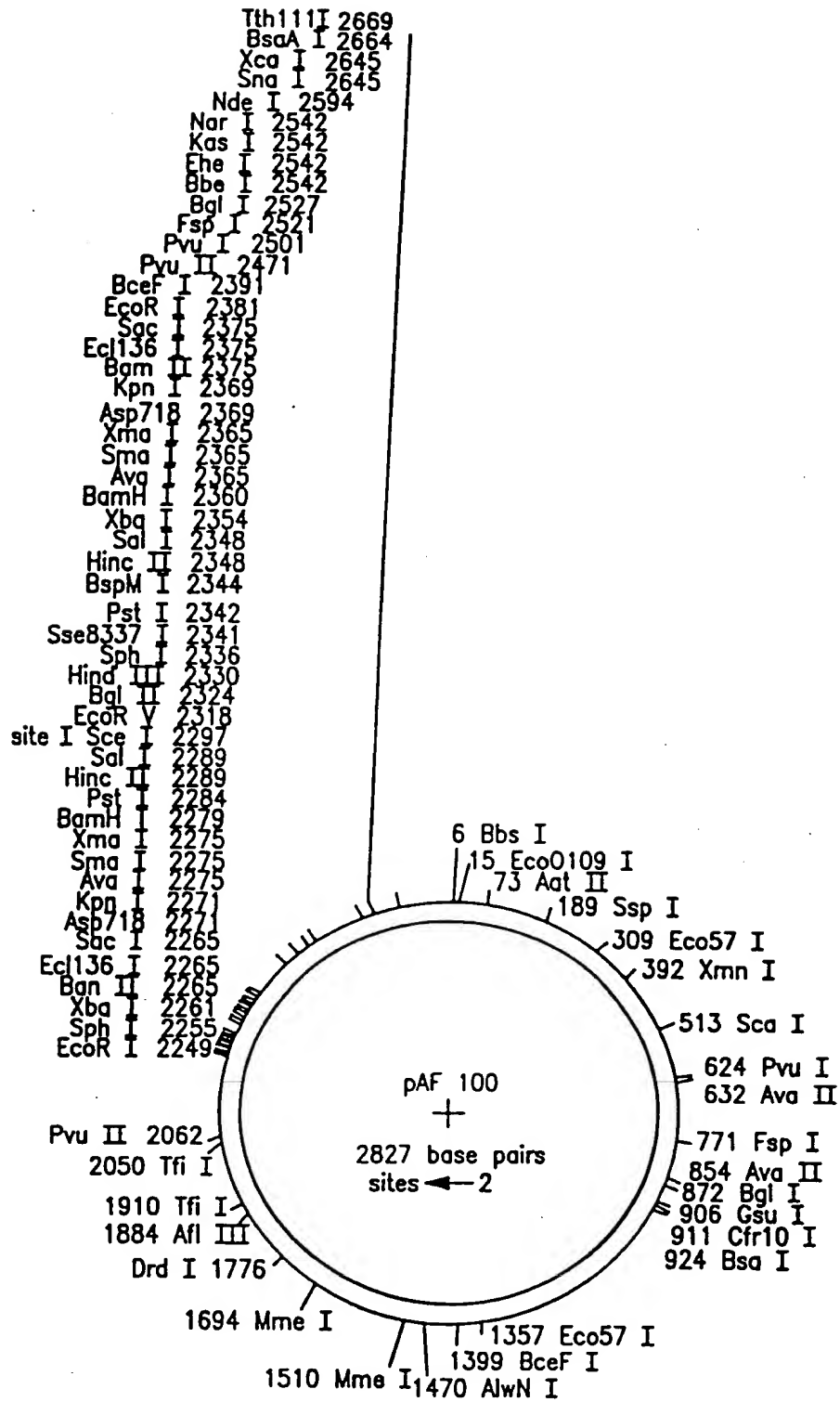


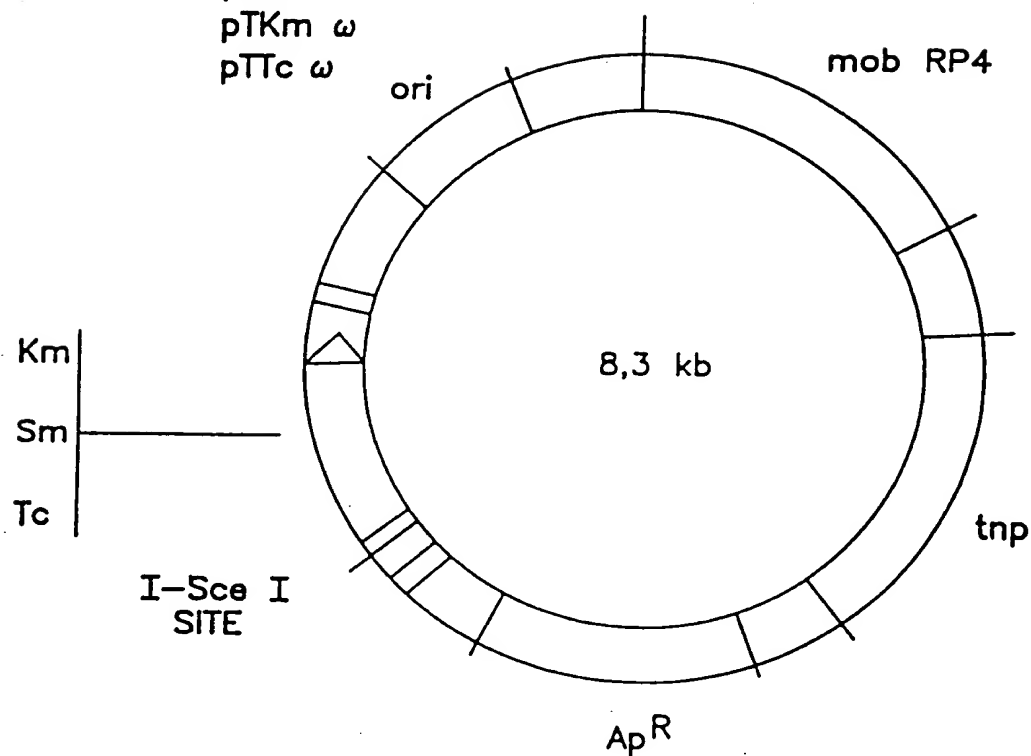
FIG. 9

Sau3A I
 Mbo I
 Dpn II
 ScrF I
 Nci I
 Msp I
 Hpa II
 Dsa V
 BstK I
 Xma I
 Sma I
 ScrF I
 Nci I
 Dsa V
 Rsa I Dpn I
 Nla IV Nla IV
 Sac I Csp6 I Alw I
 HgiA I BstK I Taq I
 Ec1136 I BsaJ I Sal I
 Sph I Rma I Kpn I BstY I Hinc II
 NspC I Bsp1286 I Bcn I Sfe I
 EcoR I Xba I Ban I BamH I Acc I I-Sce I
 Apo I Nla III Ban II Bcn I Sfc I Rma I
 Taq I Nsp7524 I Alu I Ava I Pst I Hga I
 Alu I Nsp I Bfa I Asp718 Alw I BspW I Bfa I EcoR V
 CCAAGCTCGAATTCGCATGCTCTAGAGCTCGGTACCCGGGATCCTGCAGTCGACGGTAGGGATAACAGGGTAATACAGAT 2320
 GGTTTCGAGCTTAAGCGTACGAGATCTCGAGCCATGGGCCCTAGGACGTCAGCTGCCATCCCTATTGTCCCATTATGTCTA 2318
 2244 2255 2262 2271 2279 2286 2296
 2247 2255 2266 2275 2284 2292
 2249 2256 2265 2275 2284 2296
 2249 2261 2271 2279 2289 2297
 2255 2265 2276 2284
 2255 2262 2271 2279 2289 2289 2290
 2265 2272 2280
 2271 2279 2280
 2272 2280
 2275 2275 2275 2275 2275 2276 2276 2276 2276 2276 2280 2280 2280

FIG. 10A

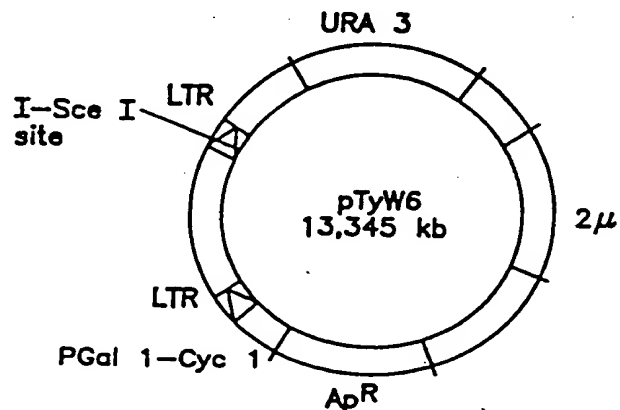
FIG. 10B

2430

[illegible]

Construction: pGP 704 from De Lorenzo, with transposase gene and insertion of the linker[I-SceI] in NotI unique site

FIG. 11



Construction: pD 123, from J.D. Boeke
with insertion of a linker[I-SceI-NotI] in BamHI

FIG. 12

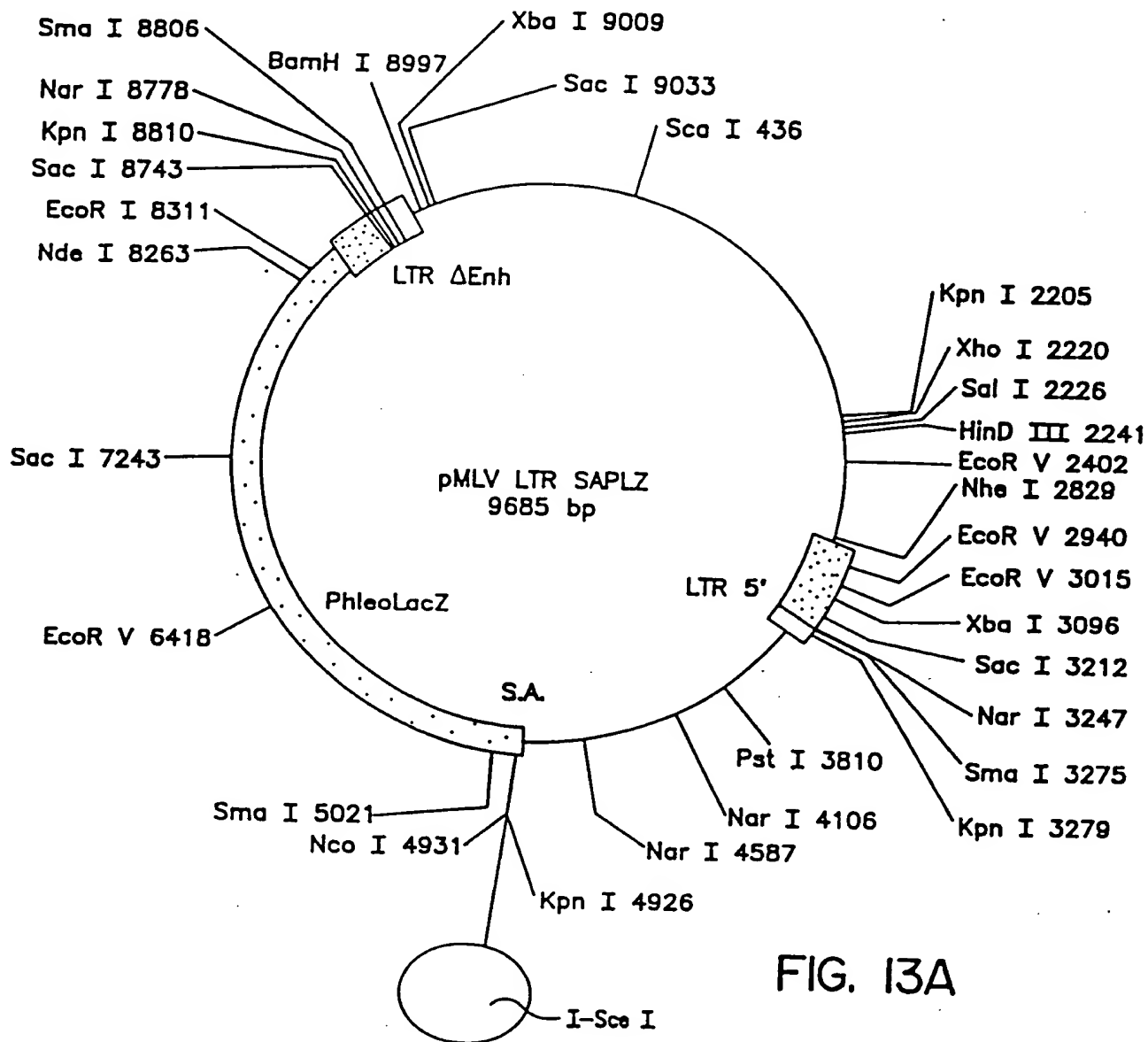


FIG. 13A

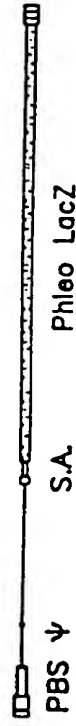
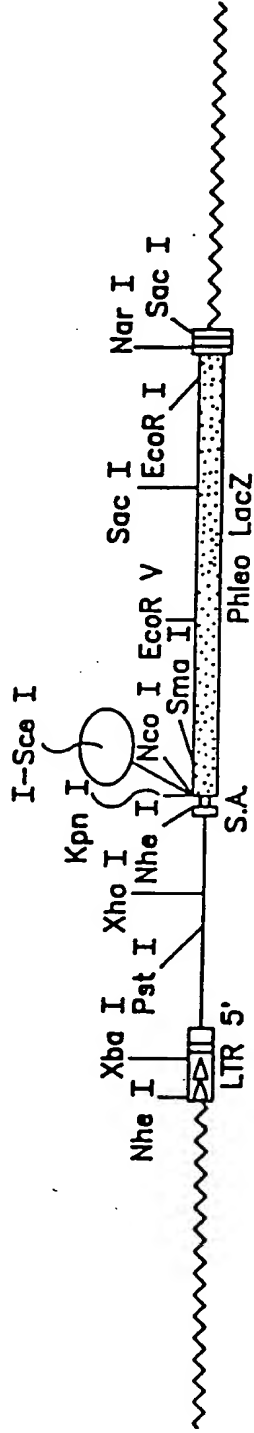


FIG. 13B

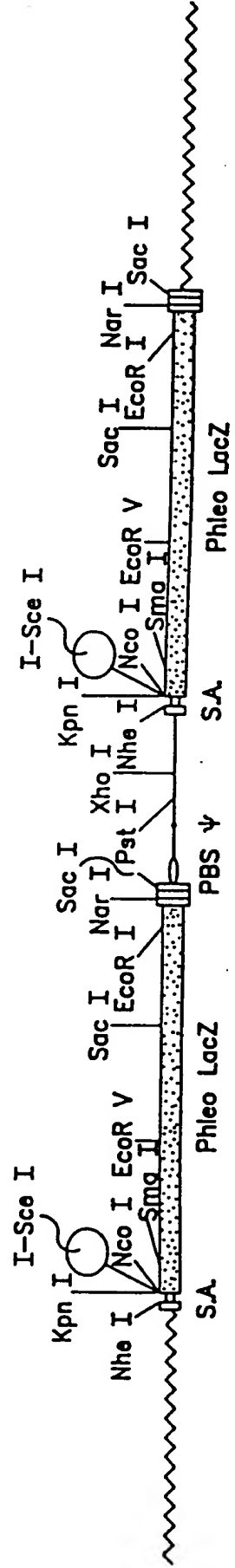


FIG. 13C

CONTROL D304 G41 M57 CONTROL
A302 E40 H81 T62

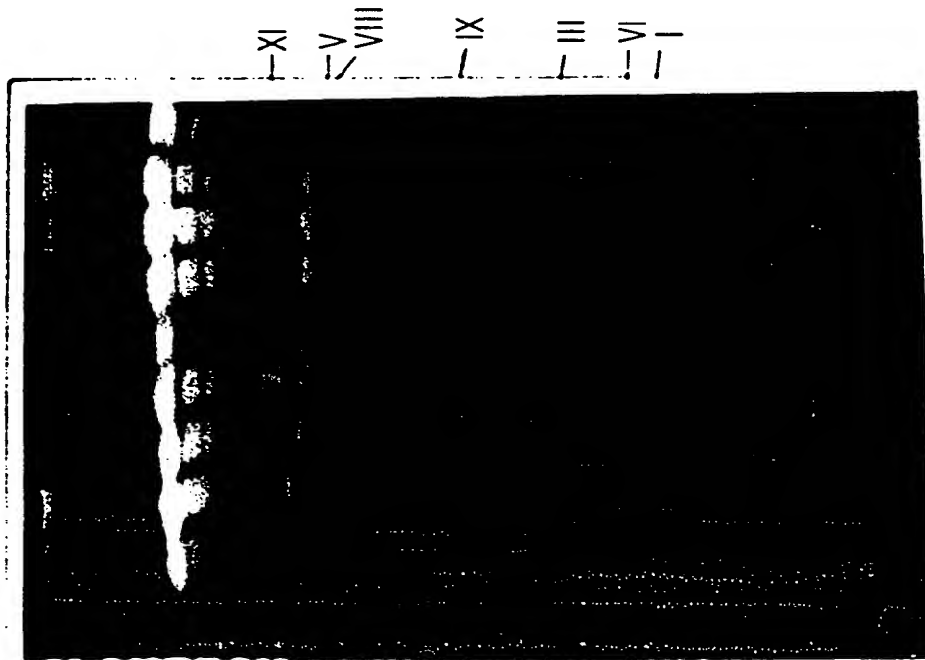
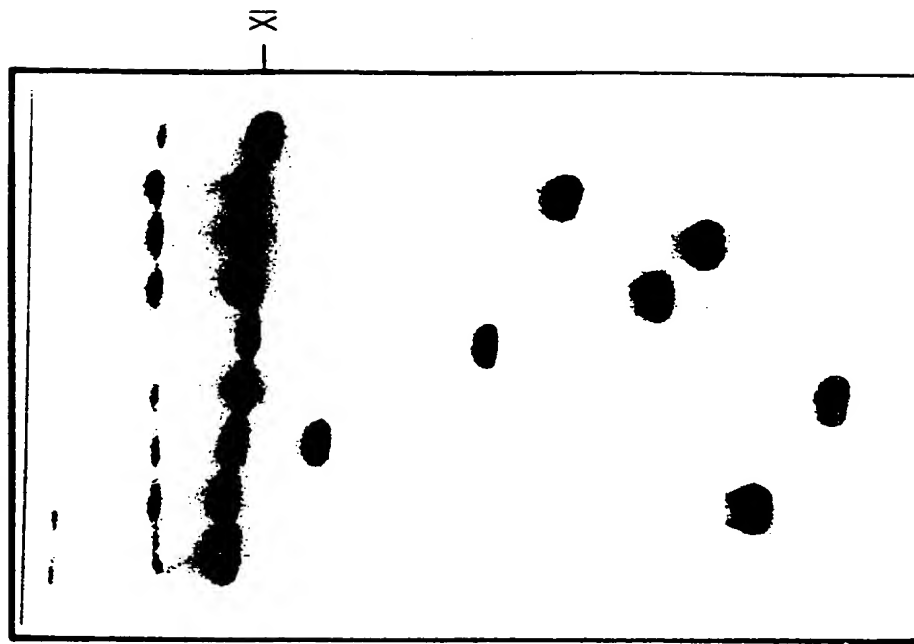


FIG. 14A

CONTROL A304 G41 M57 CONTROL
A302 E40 H81 T62



LEFT END PROBE
COSMID pUKG 040

FIG. 14B

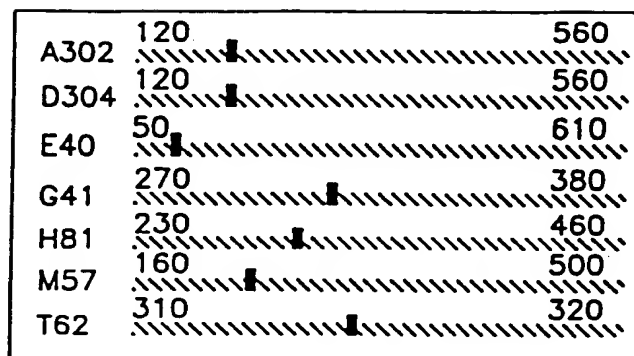


FIG. 15A

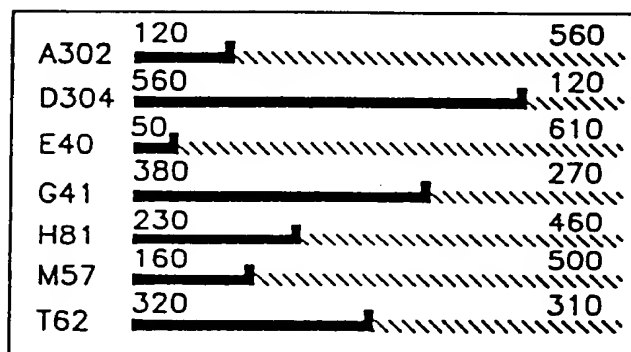


FIG. 15B

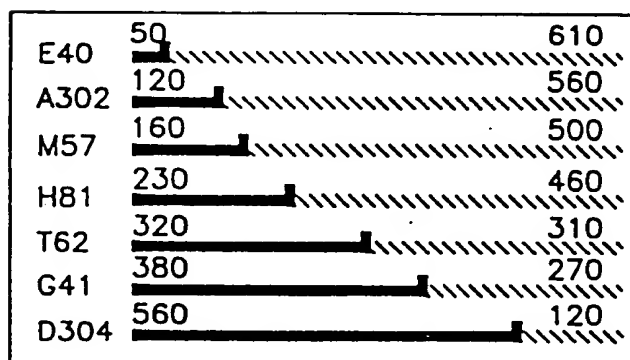


FIG. 15C

FIG. 15D

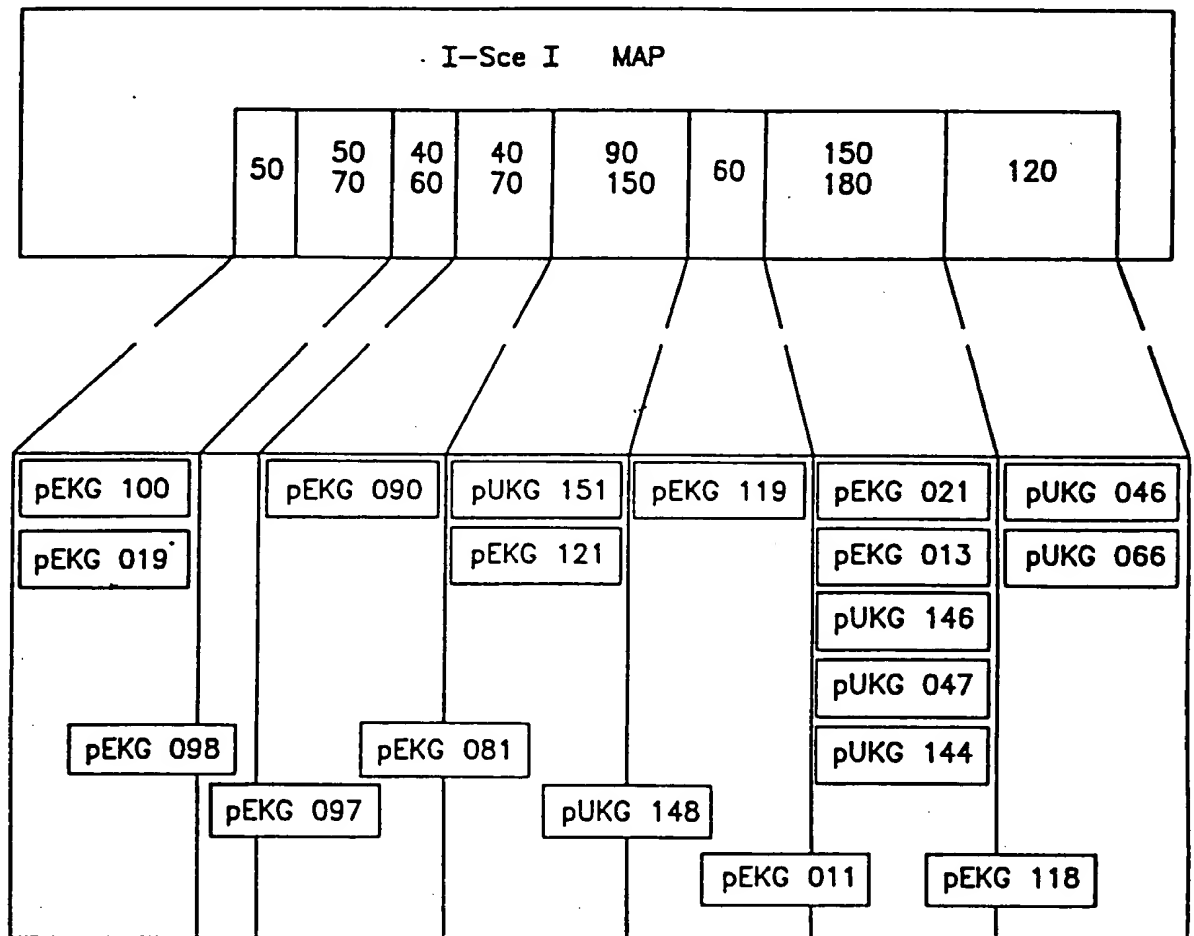


FIG. 15E

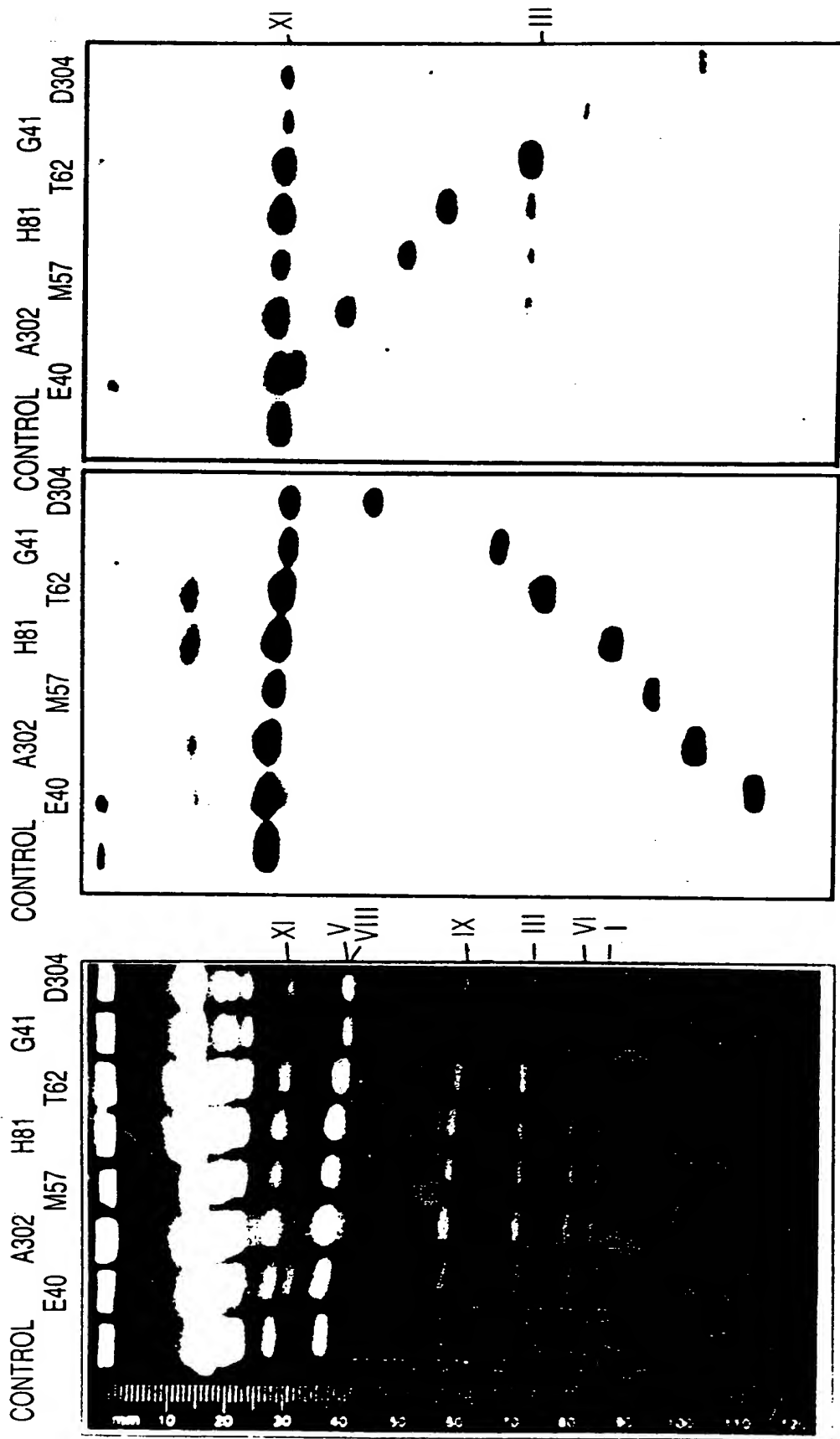


FIG. 16A

FIG. 16B

FIG. 16C

LAD- pEKG pEKG pEKG pEKG pEKG pEKG pUKG pUKG pUKG pUKG
 DER 019 097 081 121 119 021 146 144 046
 pEKG pEKG pUKG pUKG pEKG pEKG pUKG pEKG pUKG
 100 098 090 151 148 011 047 013 118 066

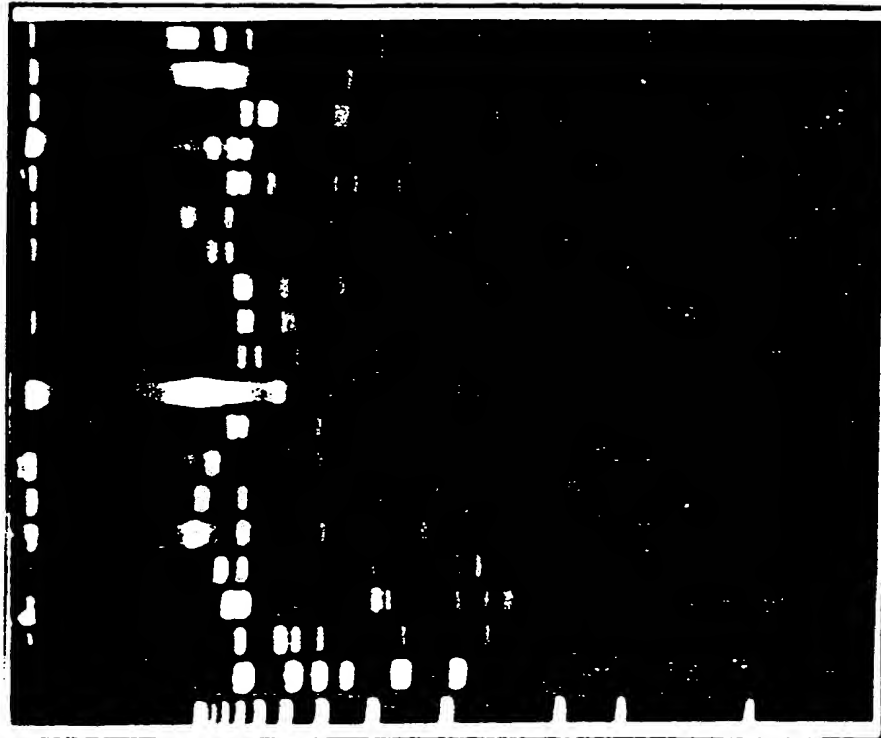


FIG. 17A

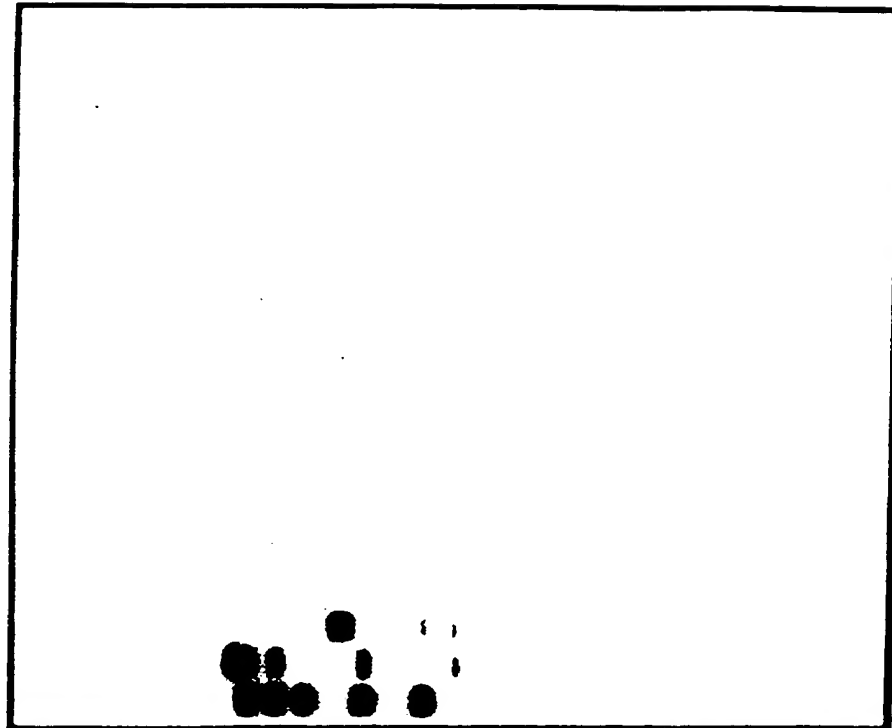


FIG. 17B

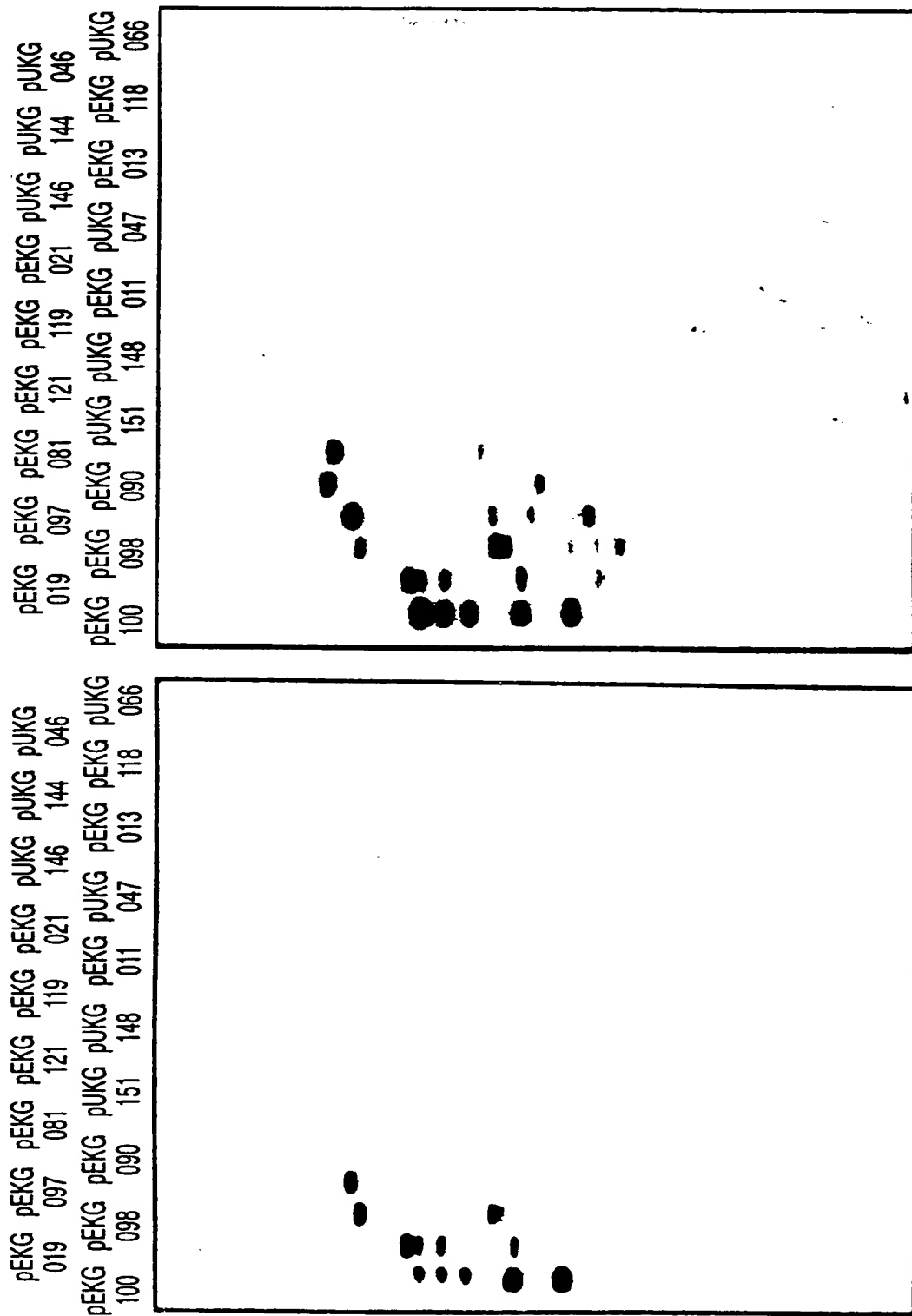


FIG. 17C

FIG. 17D

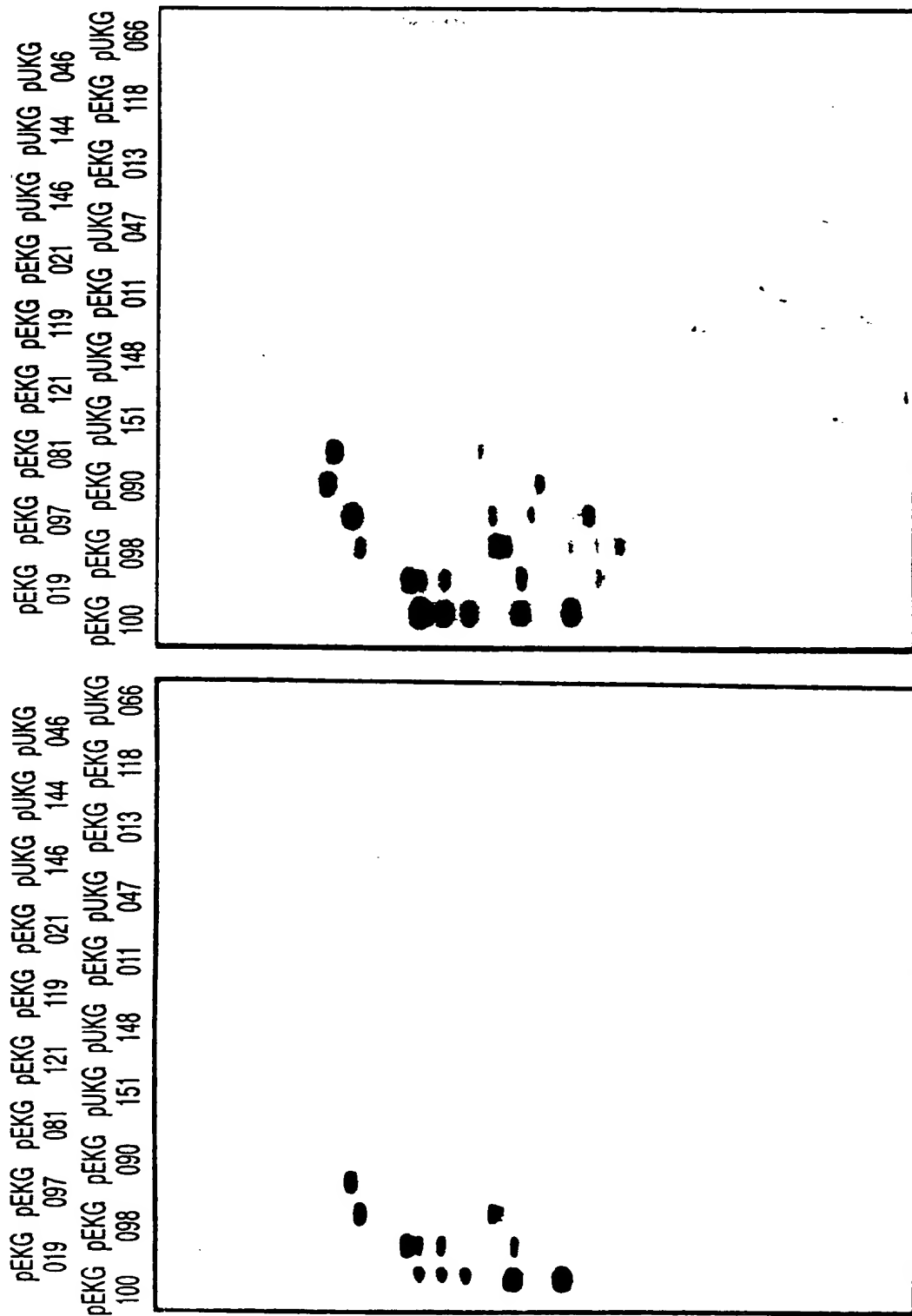


FIG. 17C

FIG. 17D

pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 019 097 081 121 119 021 146 144 046
 pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 100 098 090 151 148 011 047 013 118 066

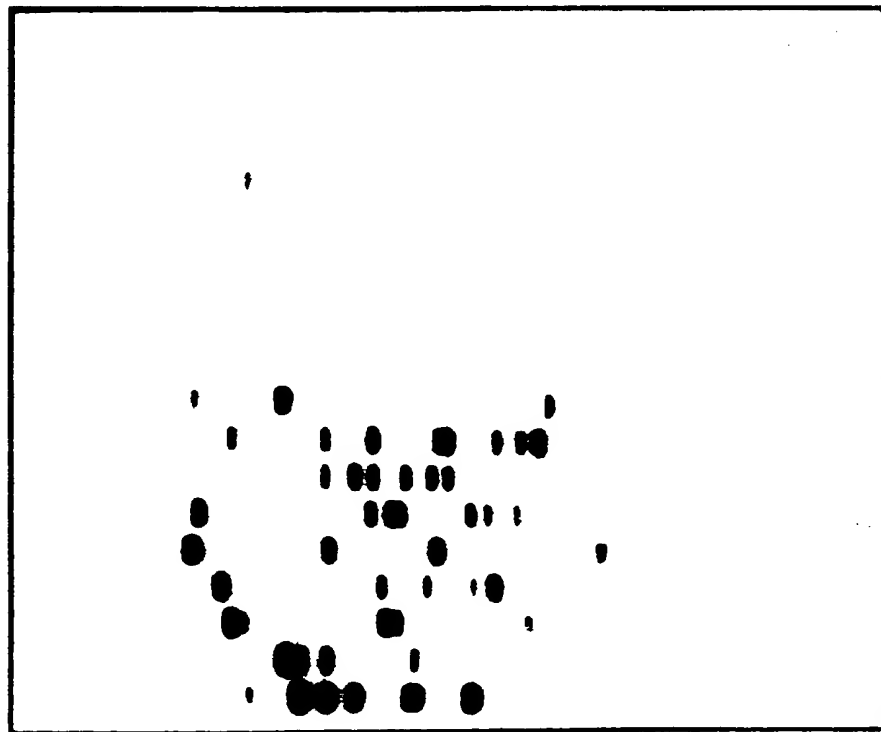


FIG. 17E

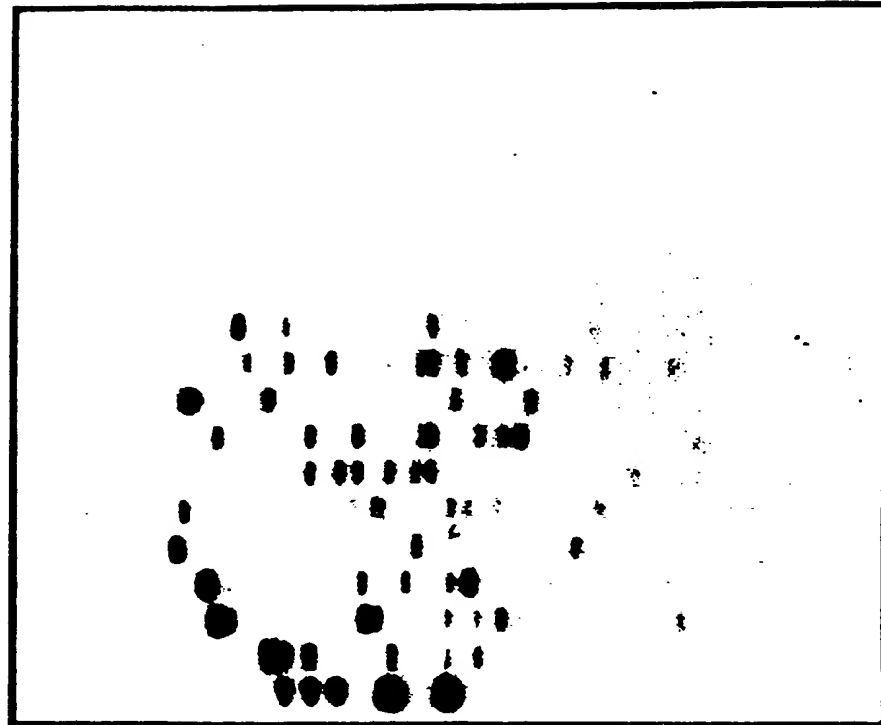


FIG. 17F

pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 019 097 081 121 119 021 146 144 046
 pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 100 098 090 151 148 011 047 013 118 066

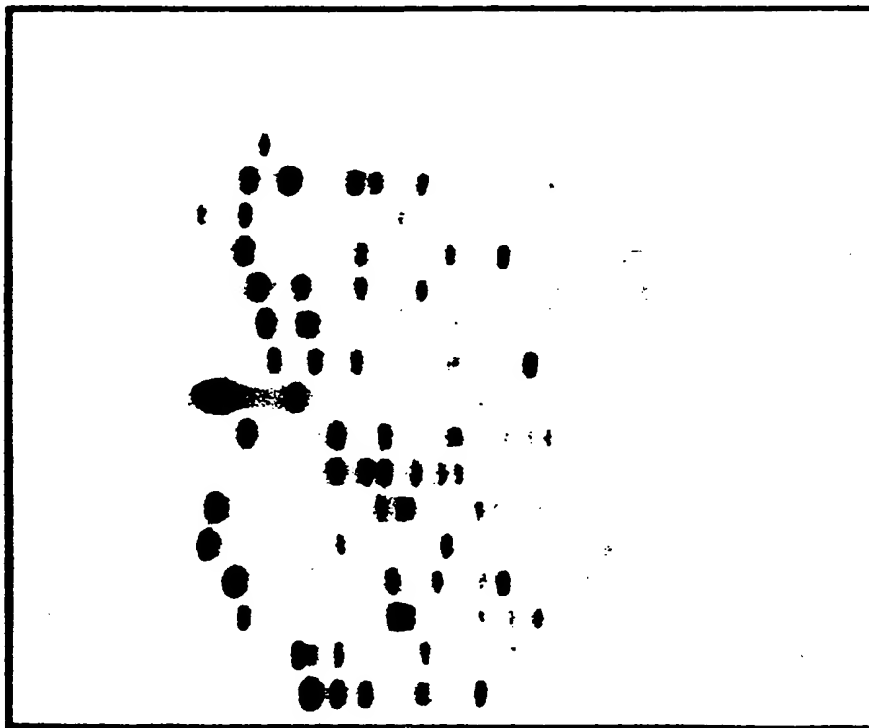


FIG. 17G

pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 019 097 081 121 119 021 146 144 046
 pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG
 100 098 090 151 148 011 047 013 118 066

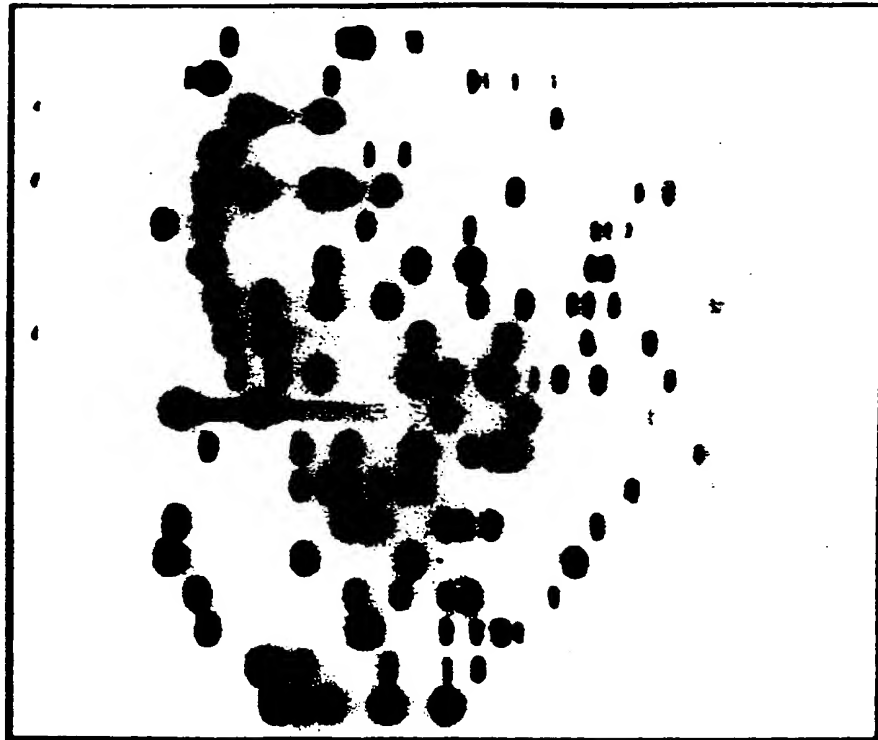


FIG. 17H

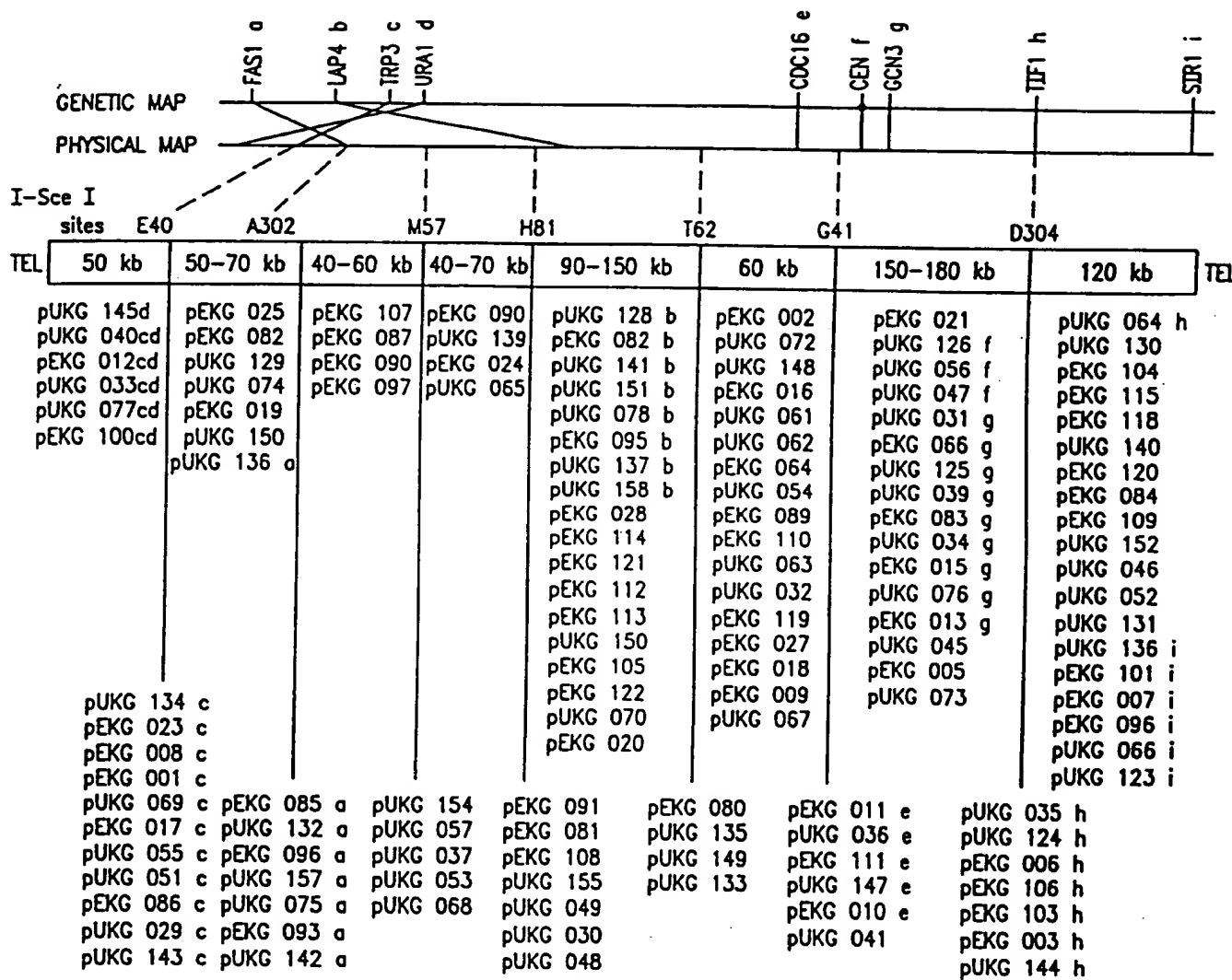


FIG. 18

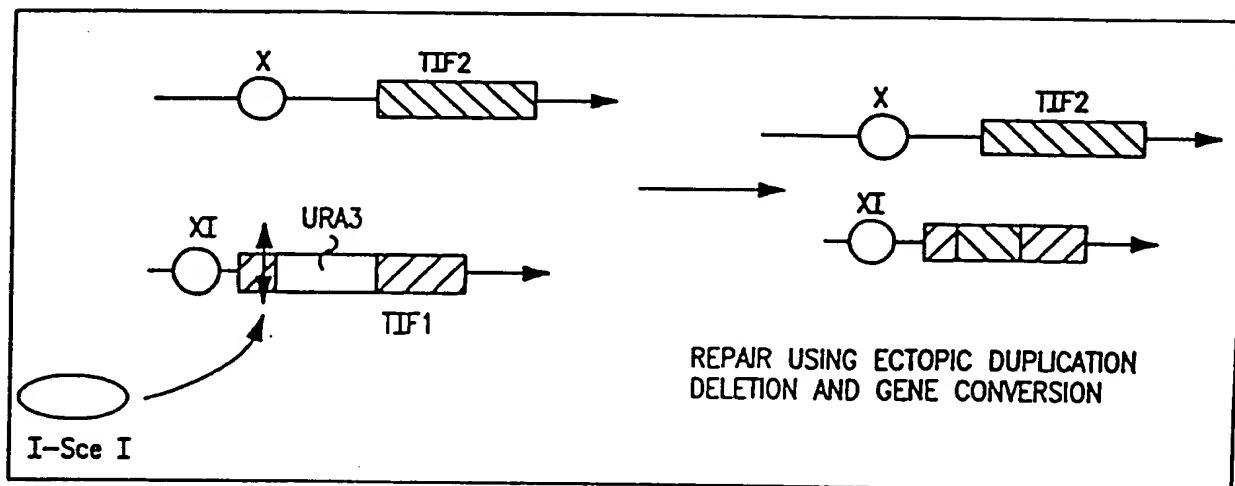


FIG. 19A

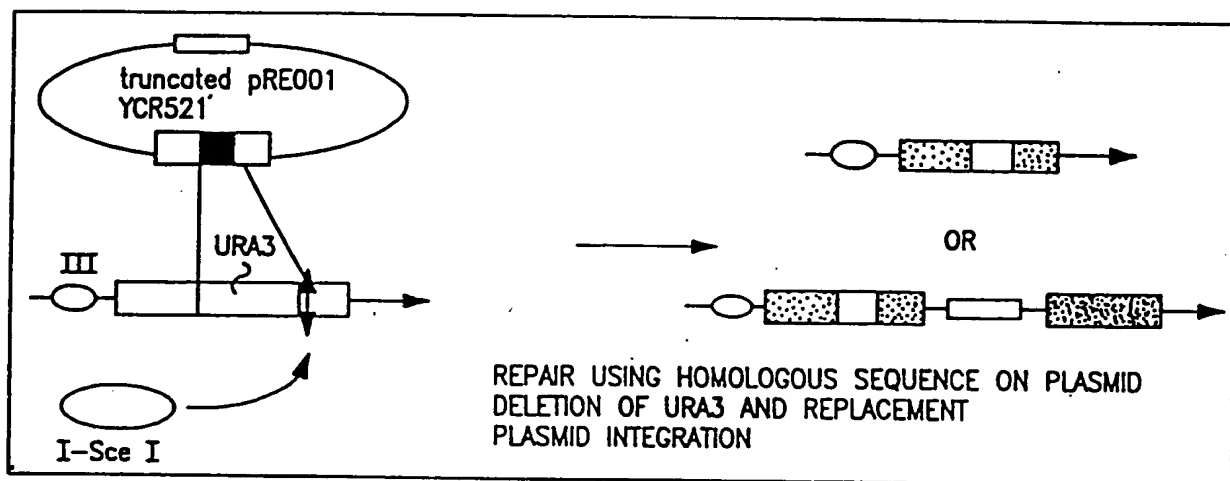


FIG. 19B